



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>A61N 43/04, C12Q 1/68, C12N 15/00, C07H 21/02, A61K 39/00</b>		A1	(11) International Publication Number: <b>WO 97/47358</b> (43) International Publication Date: 18 December 1997 (18.12.97)
(21) International Application Number: <b>PCT/US97/09884</b> (22) International Filing Date: 6 June 1997 (06.06.97)		(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, HU, IL, IS, JP, KG, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TR, TT, UA, US, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(30) Priority Data: 60/020,494 11 June 1996 (11.06.96) US 9614731.9 12 July 1996 (12.07.96) GB 60/033,534 20 December 1996 (20.12.96) US		Published With international search report.	
(71) Applicant (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): DONNELLY, John, J. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). LU, Tong-Ming [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). LIU, Margaret, A. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).			
(74) Common Representative: MERCK & CO., INC; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).			
<p><b>(54) Title:</b> SYNTHETIC HEPATITIS C GENES</p> <p><b>(57) Abstract</b></p> <p>This invention relates to novel methods and formulations of nucleic acid pharmaceutical products, specifically formulations of nucleic acid vaccine products and nucleic acid gene therapy products.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IR	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KR	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		

- 1 -

**TITLE OF THE INVENTION**  
**SYNTHETIC HEPATITIS C GENES**

**CROSS-REFERENCE TO RELATED APPLICATIONS**

5       Not applicable.

**STATEMENT REGARDING FEDERALLY-SPONSORED R&D**

Not applicable.

**10   REFERENCE TO MICROFICHE APPENDIX**

Not applicable.

**FIELD OF THE INVENTION**

Not applicable.

15

**BACKGROUND OF THE INVENTION**

This invention relates to novel nucleic acid pharmaceutical products, specifically nucleic acid vaccine products. The nucleic acid vaccine products, when introduced directly into muscle cells, induce the 20 production of immune responses which specifically recognize Hepatitis C virus (HCV).

**Hepatitis C Virus**

Non-A, Non-B hepatitis (NANBH) is a transmissible disease 25 (or family of diseases) that is believed to be virally induced, and is distinguishable from other forms of virus-associated liver disease, such as those caused by hepatitis A virus (HAV), hepatitis B virus (HBV), delta hepatitis virus (HDV), cytomegalovirus (CMV) or Epstein-Barr virus (EBV). Epidemiologic evidence suggests that there may be three 30 types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the sporadically occurring (community acquired) type. However, the number of causative agents is unknown. Recently, a new viral species, hepatitis C virus (HCV) has been identified as the primary (if not only) cause of blood-associated NANBH (BB-NANBH).

- 2 -

Hepatitis C appears to be the major form of transfusion-associated hepatitis in a number of countries, including the United States and Japan. There is also evidence implicating HCV in induction of hepatocellular carcinoma. Thus, a need exists for an effective method 5 for preventing or treating HCV infection: currently, there is none.

The HCV may be distantly related to the flaviviridae. The Flavivirus family contains a large number of viruses which are small, enveloped pathogens of man. The morphology and composition of Flavivirus particles are known, and are discussed in M. A. Brinton, in 10 "The Viruses: The Togaviridae And Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press, 1986), pp. 327-374. Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm. 15 Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections measuring about 5-10 nm in length with terminal knobs about 2 nm in diameter. Typical examples of the family include Yellow Fever virus, West Nile virus, and Dengue Fever virus. They possess positive-stranded RNA genomes (about 11,000 20 nucleotides) that are slightly larger than that of HCV and encode a polyprotein precursor of about 3500 amino acids. Individual viral proteins are cleaved from this precursor polypeptide.

The genome of HCV appears to be single-stranded RNA containing about 10,000 nucleotides. The genome is positive-stranded, 25 and possesses a continuous translational open reading frame (ORF) that encodes a polyprotein of about 3,000 amino acids. In the ORF, the structural proteins appear to be encoded in approximately the first quarter of the N-terminal region, with the majority of the polyprotein attributed to non-structural proteins. When compared with all known 30 viral sequences, small but significant co-linear homologies are observed with the nonstructural proteins of the Flavivirus family, and with the pestiviruses (which are now also considered to be part of the Flavivirus family).

- 3 -

Intramuscular inoculation of polynucleotide constructs, i.e., DNA plasmids encoding proteins have been shown to result in the in situ generation of the protein in muscle cells. By using cDNA plasmids encoding viral proteins, both antibody and CTL responses were 5 generated, providing homologous and heterologous protection against subsequent challenge with either the homologous or cross-strain protection, respectively. Each of these types of immune responses offers a potential advantage over existing vaccination strategies. The use of PNVs (polynucleotide vaccines) to generate antibodies may result 10 in an increased duration of the antibody responses as well as the provision of an antigen that can have both the exact sequence of the clinically circulating strain of virus as well as the proper post-translational modifications and conformation of the native protein (vs. a recombinant protein). The generation of CTL responses by this means 15 offers the benefits of cross-strain protection without the use of a live potentially pathogenic vector or attenuated virus.

Therefore, this invention contemplates methods for introducing nucleic acids into living tissue to induce expression of proteins. The invention provides a method for introducing viral 20 proteins into the antigen processing pathway to generate virus-specific immune responses including, but not limited to, CTLs. Thus, the need for specific therapeutic agents capable of eliciting desired prophylactic immune responses against viral pathogens is met for HCV virus by this invention. Of particular importance in this therapeutic approach is the 25 ability to induce T-cell immune responses which can prevent infections even of virus strains which are heterologous to the strain from which the antigen gene was obtained. Therefore, this invention provides DNA constructs encoding viral proteins of the hepatitis C virus core, envelope (E1), nonstructural (NS5) genes or any other HCV genes which encode 30 products which generate specific immune responses including but not limited to CTLs.

DNA Vaccines

Benvenisty, N., and Reshef, L. [PNAS 83, 9551-9555, (1986)] showed that CaCl<sub>2</sub>-precipitated DNA introduced into mice intraperitoneally (i.p.), intravenously (i.v.) or intramuscularly (i.m.) 5 could be expressed. The i.m. injection of DNA expression vectors without CaCl<sub>2</sub> treatment in mice resulted in the uptake of DNA by the muscle cells and expression of the protein encoded by the DNA. The plasmids were maintained episomally and did not replicate. Subsequently, persistent expression has been observed after i.m. 10 injection in skeletal muscle of rats, fish and primates, and cardiac muscle of rats. The technique of using nucleic acids as therapeutic agents was reported in WO90/11092 (4 October 1990), in which polynucleotides were used to vaccinate vertebrates.

It is not necessary for the success of the method that 15 immunization be intramuscular. The introduction of gold microprojectiles coated with DNA encoding bovine growth hormone (BGH) into the skin of mice resulted in production of anti-BGH antibodies in the mice. A jet injector has been used to transfect skin, muscle, fat, and mammary tissues of living animals. Various methods 20 for introducing nucleic acids have been reviewed. Intravenous injection of a DNA:cationic liposome complex in mice was shown by Zhu et al., [Science 261:209-211 (9 July 1993) to result in systemic expression of a cloned transgene. Ulmer et al., [Science 259:1745-1749, (1993)] 25 reported on the heterologous protection against influenza virus infection by intramuscular injection of DNA encoding influenza virus proteins.

The need for specific therapeutic and prophylactic agents capable of eliciting desired immune responses against pathogens and tumor antigens is met by the instant invention. Of particular importance in this therapeutic approach is the ability to induce T-cell 30 immune responses which can prevent infections or disease caused even by virus strains which are heterologous to the strain from which the antigen gene was obtained. This is of particular concern when dealing with HIV as this virus has been recognized to mutate rapidly and many virulent isolates have been identified [see, for example, LaRosa et al.,

- 5 -

Science 249:932-935 (1990), identifying 245 separate HIV isolates]. In response to this recognized diversity, researchers have attempted to generate CTLs based on peptide immunization. Thus, Takahashi et al., [Science 255:333-336 (1992)] reported on the induction of broadly 5 cross-reactive cytotoxic T cells recognizing an HIV envelope (gp160) determinant. However, those workers recognized the difficulty in achieving a truly cross-reactive CTL response and suggested that there is a dichotomy between the priming or restimulation of T cells, which is very stringent, and the elicitation of effector function, including 10 cytotoxicity, from already stimulated CTLs.

Wang et al. reported on elicitation of immune responses in mice against HIV by intramuscular inoculation with a cloned, genomic (unspliced) HIV gene. However, the level of immune responses achieved in these studies was very low. In addition, the Wang et al., 15 DNA construct utilized an essentially genomic piece of HIV encoding contiguous Tat/REV-gp160-Tat/REV coding sequences. As is described in detail below, this is a suboptimal system for obtaining high-level expression of the gp160. It also is potentially dangerous because expression of Tat contributes to the progression of Karposi's Sarcoma.

WO 93/17706 describes a method for vaccinating an animal 20 against a virus, wherein carrier particles were coated with a gene construct and the coated particles are accelerated into cells of an animal.

The instant invention contemplates any of the known 25 methods for introducing polynucleotides into living tissue to induce expression of proteins. However, this invention provides a novel immunogen for introducing proteins into the antigen processing pathway to efficiently generate specific CTLs and antibodies.

#### Codon Usage and Codon Context

The codon pairings of organisms are highly nonrandom, 30 and differ from organism to organism. This information is used to construct and express altered or synthetic genes having desired levels of translational efficiency, to determine which regions in a genome are protein coding regions, to introduce translational pause sites into

- 6 -

heterologous genes, and to ascertain relationship or ancestral origin of nucleotide sequences

The expression of foreign heterologous genes in transformed organisms is now commonplace. A large number of 5 mammalian genes, including, for example, murine and human genes, have been successfully inserted into single celled organisms. Standard techniques in this regard include introduction of the foreign gene to be expressed into a vector such as a plasmid or a phage and utilizing that vector to insert the gene into an organism. The native promoters for 10 such genes are commonly replaced with strong promoters compatible with the host into which the gene is inserted. Protein sequencing machinery permits elucidation of the amino acid sequences of even minute quantities of native protein. From these amino acid sequences, DNA sequences coding for those proteins can be inferred. DNA 15 synthesis is also a rapidly developing art, and synthetic genes corresponding to those inferred DNA sequences can be readily constructed.

Despite the burgeoning knowledge of expression systems and recombinant DNA, significant obstacles remain when one attempts 20 to express a foreign or synthetic gene in an organism. Many native, active proteins, for example, are glycosylated in a manner different from that which occurs when they are expressed in a foreign host. For this reason, eukaryotic hosts such as yeast may be preferred to bacterial hosts for expressing many mammalian genes. The glycosylation 25 problem is the subject of continuing research.

Another problem is more poorly understood. Often 30 translation of a synthetic gene, even when coupled with a strong promoter, proceeds much less efficiently than would be expected. The same is frequently true of exogenous genes foreign to the expression organism. Even when the gene is transcribed in a sufficiently efficient manner that recoverable quantities of the translation product are produced, the protein is often inactive or otherwise different in properties from the native protein.

- 7 -

It is recognized that the latter problem is commonly due to differences in protein folding in various organisms. The solution to this problem has been elusive, and the mechanisms controlling protein folding are poorly understood.

5        The problems related to translational efficiency are believed to be related to codon context effects. The protein coding regions of genes in all organisms are subject to a wide variety of functional constraints, some of which depend on the requirement for encoding a properly functioning protein, as well as appropriate  
10      translational start and stop signals. However, several features of protein coding regions have been discerned which are not readily understood in terms of these constraints. Two important classes of such features are those involving codon usage and codon context.

15      It is known that codon utilization is highly biased and varies considerably between different organisms. Codon usage patterns have been shown to be related to the relative abundance of tRNA isoacceptors. Genes encoding proteins of high versus low abundance show differences in their codon preferences. The possibility that biases in codon usage alter peptide elongation rates has been widely discussed.  
20      While differences in codon use are associated with differences in translation rates, direct effects of codon choice on translation have been difficult to demonstrate. Other proposed constraints on codon usage patterns include maximizing the fidelity of translation and optimizing the kinetic efficiency of protein synthesis.

25      Apart from the non-random use of codons, considerable evidence has accumulated that codon/anticodon recognition is influenced by sequences outside the codon itself, a phenomenon termed "codon context." There exists a strong influence of nearby nucleotides on the efficiency of suppression of nonsense codons as well as missense codons.  
30      Clearly, the abundance of suppressor activity in natural bacterial populations, as well as the use of "termination" codons to encode selenocysteine and phosphoserine require that termination be context-dependent. Similar context effects have been shown to influence the fidelity of translation, as well as the efficiency of translation initiation.

Statistical analyses of protein coding regions of E. coli have demonstrate another manifestation of "codon context." The presence of a particular codon at one position strongly influences the frequency of occurrence of certain nucleotides in neighboring codons, and these 5 context constraints differ markedly for genes expressed at high versus low levels. Although the context effect has been recognized, the predictive value of the statistical rules relating to preferred nucleotides adjacent to codons is relatively low. This has limited the utility of such nucleotide preference data for selecting codons to effect desired levels 10 of translational efficiency.

The advent of automated nucleotide sequencing equipment has made available large quantities of sequence data for a wide variety of organisms. Understanding those data presents substantial difficulties. For example, it is important to identify the coding regions of the 15 genome in order to relate the genetic sequence data to protein sequences. In addition, the ancestry of the genome of certain organisms is of substantial interest. It is known that genomes of some organisms are of mixed ancestry. Some sequences that are viral in origin are now stably incorporated into the genome of eukaryotic organisms. The viral 20 sequences themselves may have originated in another substantially unrelated species. An understanding of the ancestry of a gene can be important in drawing proper analogies between related genes and their translation products in other organisms.

There is a need for a better understanding of codon context 25 effects on translation, and for a method for determining the appropriate codons for any desired translational effect. There is also a need for a method for identifying coding regions of the genome from nucleotide sequence data. There is also a need for a method for controlling protein folding and for insuring that a foreign gene will fold appropriately 30 when expressed in a host. Genes altered or constructed in accordance with desired translational efficiencies would be of significant worth.

Another aspect of the practice of recombinant DNA techniques for the expression by microorganisms of proteins of industrial and pharmaceutical interest is the phenomenon of "codon

- 9 -

preference". While it was earlier noted that the existing machinery for gene expression is genetically transformed host cells will "operate" to construct a given desired product, levels of expression attained in a microorganism can be subject to wide variation, depending in part on

5 specific alternative forms of the amino acid-specifying genetic code present in an inserted exogenous gene. A "triplet" codon of four possible nucleotide bases can exist in 64 variant forms. That these forms provide the message for only 20 different amino acids (as well as transcription initiation and termination) means that some amino acids

10 can be coded for by more than one codon. Indeed, some amino acids have as many as six "redundant", alternative codons while some others have a single, required codon. For reasons not completely understood, alternative codons are not at all uniformly present in the endogenous DNA of differing types of cells and there appears to exist a variable

15 natural hierarchy or "preference" for certain codons in certain types of cells.

As one example, the amino acid leucine is specified by any of six DNA codons including CTA, CTC, CTG, CTT, TTA, and TTG (which correspond, respectively, to the mRNA codons, CUA, CUC, CUG, CUU, UUA and UUG). Exhaustive analysis of genome codon frequencies for microorganisms has revealed endogenous DNA of E. coli most commonly contains the CTG leucine-specifying codon, while the DNA of yeasts and slime molds most commonly includes a TTA leucine-specifying codon. In view of this hierarchy, it is generally held

20 that the likelihood of obtaining high levels of expression of a leucine-rich polypeptide by an E. coli host will depend to some extent on the frequency of codon use. For example, a gene rich in TTA codons will in all probability be poorly expressed in E. coli, whereas a CTG rich gene will probably highly express the polypeptide. Similarly, when

25 yeast cells are the projected transformation host cells for expression of a leucine-rich polypeptide, a preferred codon for use in an inserted DNA would be TTA.

The implications of codon preference phenomena on recombinant DNA techniques are manifest, and the phenomenon may

- 10 -

serve to explain many prior failures to achieve high expression levels of exogenous genes in successfully transformed host organisms-a less "preferred" codon may be repeatedly present in the inserted gene and the host cell machinery for expression may not operate as efficiently.

5 This phenomenon suggests that synthetic genes which have been designed to include a projected host cell's preferred codons provide a preferred form of foreign genetic material for practice of recombinant DNA techniques.

10 **Protein Trafficking**

The diversity of function that typifies eukaryotic cells depends upon the structural differentiation of their membrane boundaries. To generate and maintain these structures, proteins must be transported from their site of synthesis in the endoplasmic reticulum to 15 predetermined destinations throughout the cell. This requires that the trafficking proteins display sorting signals that are recognized by the molecular machinery responsible for route selection located at the access points to the main trafficking pathways. Sorting decisions for most proteins need to be made only once as they traverse their biosynthetic pathways since their final destination, the cellular location 20 at which they perform their function, becomes their permanent residence.

Maintenance of intracellular integrity depends in part on the selective sorting and accurate transport of proteins to their correct 25 destinations. Over the past few years the dissection of the molecular machinery for targeting and localization of proteins has been studied vigorously. Defined sequence motifs have been identified on proteins which can act as 'address labels'. A number of sorting signals have been found associated with the cytoplasmic domains of membrane proteins.

30

**SUMMARY OF THE INVENTION**

This invention relates to novel formulations of nucleic acid pharmaceutical products, specifically nucleic acid vaccine products. The nucleic acid products, when introduced directly into muscle cells,

- 11 -

induce the production of immune responses which specifically recognize Hepatitis C virus (HCV).

## BRIEF DESCRIPTION OF THE DRAWINGS

5      Figure 1 shows the nucleotide sequence of the V1Ra vector.  
Figure 2 is a diagram of the V1Ra vector.  
Figure 3 is a diagram of the Vtpa vector.  
Figure 4 is the VUb vector  
Figure 5 shows an optimized sequence of the HCV core  
10     antigen.  
Figure 6 shows V1Ra.HCV1CorePAb, Vtpa.HCV1CorePAb  
and VUb.HCV1CorePAb.  
Figure 7 shows the Hepatitis C Virus Core Antigen  
Sequence.  
15     Figure 8 shows codon utilization in human protein-coding  
sequences (from Lathe et al.).  
Figure 9 shows an optimized sequence of the HCV E1  
protein.  
Figure 10 shows an optimized sequence of the HCV E2  
protein.  
20     Figure 11 shows an optimized sequence of the HCV E1 +E2  
proteins.  
Figure 12 shows an optimized sequence of the HCV NS5a  
protein.  
25     Figure 13 shows an optimized sequence of the HCV NS5b  
protein.

## DETAILED DESCRIPTION OF THE INVENTION

This invention relates to novel formulations of nucleic acid pharmaceutical products, specifically nucleic acid vaccine products. The nucleic acid vaccine products, when introduced directly into muscle cells, induce the production of immune responses which specifically recognize Hepatitis C virus (HCV).

- 12 -

Non-A, Non-B hepatitis (NANBH) is a transmissible disease (or family of diseases) that is believed to be virally induced, and is distinguishable from other forms of virus-associated liver disease, such as those caused by hepatitis A virus (HAV), hepatitis B virus (HBV),

5 delta hepatitis virus (HDV), cytomegalovirus (CMV) or Epstein-Barr virus (EBV). Epidemiologic evidence suggests that there may be three types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the sporadically occurring (community acquired) type. However, the number of causative agents is unknown. Recently, a

10 new viral species, hepatitis C virus (HCV) has been identified as the primary (if not only) cause of blood-associated NANBH (BB-NANBH). Hepatitis C appears to be the major form of transfusion-associated hepatitis in a number of countries, including the United States and Japan. There is also evidence implicating HCV in induction of

15 hepatocellular carcinoma. Thus, a need exists for an effective method for preventing or treating HCV infection: currently, there is none.

The HCV may be distantly related to the flaviviridae. The Flavivirus family contains a large number of viruses which are small, enveloped pathogens of man. The morphology and composition of

20 Flavivirus particles are known, and are discussed in M. A. Brinton, in "The Viruses: The Togaviridae And Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press, 1986), pp. 327-374. Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid

25 bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections measuring about 5-10 nm in length with terminal knobs about 2 nm in diameter. Typical examples of the family include Yellow Fever virus, West Nile virus, and Dengue Fever

30 virus. They possess positive-stranded RNA genomes (about 11,000 nucleotides) that are slightly larger than that of HCV and encode a polyprotein precursor of about 3500 amino acids. Individual viral proteins are cleaved from this precursor polypeptide.

The genome of HCV appears to be single-stranded RNA containing about 10,000 nucleotides. The genome is positive-stranded, and possesses a continuous translational open reading frame (ORF) that encodes a polyprotein of about 3,000 amino acids. In the ORF, the 5 structural proteins appear to be encoded in approximately the first quarter of the N-terminal region, with the majority of the polyprotein attributed to non-structural proteins. When compared with all known viral sequences, small but significant co-linear homologics are observed with the nonstructural proteins of the Flavivirus family, and with the 10 pestiviruses (which are now also considered to be part of the Flavivirus family).

Intramuscular inoculation of polynucleotide constructs, i.e., DNA plasmids encoding proteins have been shown to result in the generation of the encoded protein *in situ* in muscle cells. By using 15 cDNA plasmids encoding viral proteins, both antibody and CTL responses were generated, providing homologous and heterologous protection against subsequent challenge with either the homologous or cross-strain protection, respectively. Each of these types of immune responses offers a potential advantage over existing vaccination 20 strategies. The use of PNVs (polynucleotide vaccines) to generate antibodies may result in an increased duration of the antibody responses as well as the provision of an antigen that can have both the exact sequence of the clinically circulating strain of virus as well as the proper post-translational modifications and conformation of the native 25 protein (vs. a recombinant protein). The generation of CTL responses by this means offers the benefits of cross-strain protection without the use of a live potentially pathogenic vector or attenuated virus.

The standard techniques of molecular biology for preparing and purifying DNA constructs enable the preparation of the 30 DNA therapeutics of this invention. While standard techniques of molecular biology are therefore sufficient for the production of the products of this invention, the specific constructs disclosed herein provide novel therapeutics which surprisingly produce cross-strain

protection, a result heretofore unattainable with standard inactivated whole virus or subunit protein vaccines.

The amount of expressible DNA to be introduced to a vaccine recipient will depend on the strength of the transcriptional and 5 translational promoters used in the DNA construct, and on the immunogenicity of the expressed gene product. In general, an immunologically or prophylactically effective dose of about 1  $\mu$ g to 1 mg, and preferably about 10  $\mu$ g to 300  $\mu$ g is administered directly into muscle tissue. Subcutaneous injection, intradermal introduction, 10 impression through the skin, and other modes of administration such as intraperitoneal, intravenous, or inhalation delivery are also contemplated. It is also contemplated that booster vaccinations are to be provided.

The DNA may be naked, that is, unassociated with any 15 proteins, adjuvants or other agents which impact on the recipients immune system. In this case, it is desirable for the DNA to be in a physiologically acceptable solution, such as, but not limited to, sterile saline or sterile buffered saline. Alternatively, the DNA may be associated with surfactants, liposomes, such as lecithin liposomes or 20 other liposomes known in the art, as a DNA-liposome mixture, (see for example WO93/24640) or the DNA may be associated with an adjuvant known in the art to boost immune responses, such as a protein or other carrier. Agents which assist in the cellular uptake of DNA, such as, but not limited to, calcium ions, detergents, viral proteins and other 25 transfection facilitating agents may also be used to advantage. These agents are generally referred to as transfection facilitating agents and as pharmaceutically acceptable carriers. As used herein, the term gene refers to a segment of nucleic acid which encodes a discrete polypeptide. The term pharmaceutical, and vaccine are used interchangeably to 30 indicate compositions useful for inducing immune responses. The terms construct, and plasmid are used interchangeably. The term vector is used to indicate a DNA into which genes may be cloned for use according to the method of this invention.

The following examples are provided to further define the invention, without limiting the invention to the specifics of the examples.

5

### EXAMPLE 1

#### V1J EXPRESSION VECTORS:

V1J is derived from vectors V1 and pUC18, a commercially available plasmid. V1 was digested with SspI and EcoRI restriction enzymes producing two fragments of DNA. The smaller of these fragments, containing the CMVintA promoter and Bovine Growth Hormone (BGH) transcription termination elements which control the expression of heterologous genes, was purified from an agarose electrophoresis gel. The ends of this DNA fragment were then "blunted" using the T4 DNA polymerase enzyme in order to facilitate its ligation to another "blunt-ended" DNA fragment.

pUC18 was chosen to provide the "backbone" of the expression vector. It is known to produce high yields of plasmid, is well-characterized by sequence and function, and is of minimum size. We removed the entire *lac* operon from this vector, which was unnecessary for our purposes and may be detrimental to plasmid yields and heterologous gene expression, by partial digestion with the HaeII restriction enzyme. The remaining plasmid was purified from an agarose electrophoresis gel, blunt-ended with the T4 DNA polymerase, treated with calf intestinal alkaline phosphatase, and ligated to the CMVintA/BGH element described above. Plasmids exhibiting either of two possible orientations of the promoter elements within the pUC backbone were obtained. One of these plasmids gave much higher yields of DNA in *E. coli* and was designated V1J. This vector's structure was verified by sequence analysis of the junction regions and was subsequently demonstrated to give comparable or higher expression of heterologous genes compared with V1. The ampicillin resistance marker was replaced with the neomycin resistance marker to yield vector V1Jneo.

- 16 -

An Sfi I site was added to V1Jneo to facilitate integration studies. A commercially available 13 base pair Sfi I linker (New England BioLabs) was added at the Kpn I site within the BGH sequence of the vector. V1Jneo was linearized with Kpn I, gel purified, blunted by T4 DNA polymerase, and ligated to the blunt Sfi I linker. Clonal isolates were chosen by restriction mapping and verified by sequencing through the linker. The new vector was designated V1Jns. Expression of heterologous genes in V1Jns (with Sfi I) was comparable to expression of the same genes in V1Jneo (with Kpn I).

5 Vector V1Ra (Sequence is shown in Figure 1; map is shown in Figure 2) was derived from vector V1R, a derivative of the V1Jns vector. Multiple cloning sites (*Bgl*II, *Kpn*I, *Eco*RV, *Eco*RI, *Sall*, and *Not*I) were introduced into V1R to create the V1Ra vector to improve the convenience of subcloning. V1Ra vector derivatives containing the

10 tpa leader sequence and ubiquitin sequence were generated (Vtpa (Figure 3) and Vub (Figure 4), respectively). Expression of viral antigen from Vtpa vector will target the antigen protein into the exocytic pathway, thus producing a secretable form of the antigen proteins. These secreted proteins are likely to be captured by

15 professional antigen presenting cells, such as macrophages and dendritic cells, and processed and presented by class II molecules to activate CD4+ Th cells. They also are more likely to efficiently simulate antibody responses. Expression of viral antigen through VUb vector will produce a ubiquitin and antigen fusion protein. The uncleavable

20 ubiquitin segment (glycine to alanine change at the cleavage site, Butt et al., JBC 263:16364, 1988) will target the viral antigen to ubiquitin-associated proteasomes for rapid degradation. The resulting peptide fragments will be transported into the ER for antigen presentation by class I molecules. This modification is attempted to enhance the class I

25 molecule-restricted CTL responses against the viral antigen (Townsend et al, JEM 168:1211, 1988).

30

- 17 -

**EXAMPLE 2**  
**DESIGN AND CONSTRUCTION OF THE SYNTHETIC GENES**

**A. Design of Synthetic Gene Segments for HCV Gene Expression:**

5           Gene segments were converted to sequences having identical translated sequences (except where noted) but with alternative codon usage as defined by R. Lathe in a research article from *J. Molec. Biol.* Vol. 183, pp. 1-12 (1985) entitled "Synthetic Oligonucleotide Probes Deduced from Amino Acid Sequence Data: Theoretical and Practical Considerations". The methodology described below was based on our hypothesis that the known inability to express a gene efficiently in mammalian cells is a consequence of the overall transcript composition. Thus, using alternative codons encoding the same protein sequence may remove the constraints on HCV gene expression.

10           Inspection of the codon usage within HCV genome revealed that a high percentage of codons were among those infrequently used by highly expressed human genes. The specific codon replacement method employed may be described as follows employing data from Lathe et al.:

15           1. Identify placement of codons for proper open reading frame.

20           2. Compare wild type codon for observed frequency of use by human genes (refer to Table 3 in Lathe et al.).

25           3. If codon is not the most commonly employed, replace it with an optimal codon for high expression based on data in Table 5.

30           4. Inspect the third nucleotide of the new codon and the first nucleotide of the adjacent codon immediately 3'- of the first. If a 5'-CG-3' pairing has been created by the new codon selection, replace it with the choice indicated in Table 5.

35           5. Repeat this procedure until the entire gene segment has been replaced.

40           6. Inspect new gene sequence for undesired sequences generated by these codon replacements (e.g., "ATTAA" sequences,

- 18 -

inadvertent creation of intron splice recognition sites, unwanted restriction enzyme sites, etc.) and substitute codons that eliminate these sequences.

7. Assemble synthetic gene segments and test for  
5 improved expression.

#### B. HCV CORE ANTIGEN SEQUENCE

The consensus core sequence of HCV was adopted from a generalized core sequence reported by Bukh et al. (PNAS, 91:8239, 10 1994). This core sequence contains all the identified CTL epitopes in both human and mouse. The gene is composed of 573 nucleotides and encodes 191 amino acids. The predicted molecular weight is about 23 kDa.

The codon replacement was conducted to eliminate codons 15 which may hinder the expression of the HCV core protein in transfected mammalian cells in order to maximize the translational efficiency of DNA vaccine. Twenty three point two percent (23.2%) of nucleotide sequence (133 out of 573 nucleotides) were altered, resulting in changes of 61.3% of the codons (117 out 191 codons) in the core antigen 20 sequence. The optimized nucleotide sequence of HCV core is shown in Figure 5.

#### C. CONSTRUCTION OF THE SYNTHETIC CORE GENE

The optimized HCV core gene (Figure 5) was constructed 25 as a synthetic gene annealed from multiple synthetic oligonucleotides. To facilitate the identification and evaluation of the synthetic gene expression in cell culture and its immunogenicity in mice, a CTL epitope derived from influenza virus nucleoprotein residues 366-374 and an antibody epitope sequence derived from SV40 T antigen residues 30 684-698 were tagged to the carboxyl terminal of the core sequence (Figure 6). For clinical use it may be desired to express the core sequence without the nucleoprotein 366-374 and SV40 T 684-698 sequences. For this reason, the sequence of the two epitopes is flanked by two *Eco*RI sites which will be used to excise this fragment of

- 19 -

sequence at a later time. Thus an embodiment of the invention for clinical use could consist of the V1Ra.HCV1CorePAb, Vtpa.HCV1CorePAb, or VUb.HCV1CorePAb plasmids that had been cut with EcoRI, annealed, and ligated to yield plasmids

5 V1Ra.HCV1Core, Vtpa.HCV1Core, and VUb.HCV1Core.

The synthetic gene was built as three separate segments in three vectors, nucleotides 1 to 80 in V1Ra, nucleotides 80 to 347 (*Bst*XI site) in pUC18, and nucleotides 347 to 573 plus the two epitope sequence in pUC18. All the segments were verified by DNA

10 sequencing, and joined together in V1Ra vector.

D. HCV Gene Expression Constructs:

In each case, the junction sequences from the 5' promoter region (CMVintA) into the cloned gene is shown. The position at which 15 the junction occurs is demarcated by a "/", which does not represent any discontinuity in the sequence.

The nomenclature for these constructs follows the convention: "Vector name-HCV strain-gene".

20

**V1Ra.HCV1.CorePAb**

--IntA--AGA TCT ACC / ATG AGC--HCV.Core.--GCC / GAA TTC GCT TCC--  
PAb Sequence--TAA / ACC CGG GAA TTC TAA A / GTC GAC--BGH--

25

**Vtpa.HCV1.CorePAb**

--IntA--ATC ACC / ATG GAT--tpa leader--GAG ATC-TTC / ATG AGC--  
HCV.Core.--GCC / GAA TTC GCT TCC--PAb Sequence--TAA / ACC CGG GAA  
TTC TAA A / GTC GAC--BGH--

30

**VUb.HCV1.CorePAb.**

--IntA--AGA TCC ACC / ATG CAG--Ubiquitin--GGT GCA GAT CTG/ ATG AGC--  
HCV.Core.--GCC / GAA TTC GCT TCC--PAb Sequence--TAA / ACC CGG GAA  
TTC TAA A / GTC GAC--BGH--

- 20 -

**V1Ra.HCV1.Core**

---IntA--AGA TCT ACC / ATG AGC--HCV.Core.--GCC / TAA A / GTC GAC--  
BGH---

**5 Vtpa.HCV1.Core**

---IntA--ATC ACC / ATG GAT--tpa leader--GAG ATC-TTC / ATG AGC--  
HCV.Core.--GCC / TAA A / GTC GAC--BGH---

**VUb.HCV1.Core**

**10** ---IntA--AGA TCC ACC / ATG CAG--Ubiquitin--GGT GCA GAT CTG/ ATG AGC--  
HCV.Core.--GCC / TAA A / GTC GAC--BGH---

**E. OTHER SYNTHETIC HCV GENES**

**15** Using similar codon optimization techniques, synthetic genes encoding the HCV E1 (Figure 9), HCV E2 (Figure 10), HCV E1+E2 (Figure 11), HCV NS5a (Figure 12) and HCV NS5b (Figure 13) proteins were created.



2481	CTATTTTAA	GGTTGGGTTG	GGGCGGAA	GGAAAGGAA	GGATTTGAA	GGAAATAAGCA	GGATCTTGTG	GGATGCGGTT	2560
2561	GGTTTATTTG	CTATGGGTTG	AGGGGTTTA	ATTAGGGGG	GGGGGGGT	AGGGGGTG	TTGGGGATG	AGGGGGTTT	2640
2642	TGGGGGTTA	AAAAGGTTG	GGTTGTTG	TTTTGGGATA	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	2720
2723	AACTGGGAGG	TGGGGGAAAG	GGGAGGTT	ATTAAGGATA	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	2800
2801	TTGGGGTTG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	2880
2882	AGGTTATTTA	GGGGGGTTA	GGGAGGTT	TTTGGGGATA	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	2960
2963	CTTATGGGTTG	AGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	3040
3041	TTTGGGGAGG	GGGGGGTTA	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	3120
3121	TTTGGGGTTG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	3200
3201	GGGGGGGTG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	3280
3281	CTATTTGAA	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	3360
3361	AAATGGGAA	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	3440
3441	GGGGGGATG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	3520
3521	TTGGGGTTG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	3600
3601	ATGGGGTTG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	3680
3681	TTGGGGTTG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	3760
3761	GGGGGGATG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	3840
3841	GGGGGGATG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	3920
3921	GGGGGGATG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	4000
4001	GGGGGGATG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	4080
4081	TAATGGGAA	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	4160
4161	TATGGGTTG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	4240
4241	GGGGGGATG	GGGGGGTTG	GGGAGGTT	GGGGGGGT	GGGGGGGT	GGGGGGTG	GGGGGGATG	GGGGGGTTT	4320

25                   6. The polynucleotide of Claim 4 from which the PAb sequence has been removed.

30                   7. The plasmid vector of Claim 5 from which the PAb sequence has been removed.

35                   8. A method for inducing immune responses in a vertebrate against HCV epitopes which comprises introducing between 1 mg and 100 mg of the polynucleotide of Claim 1 into the tissue of the vertebrate.

40                   9. A method for inducing immune responses against infection or disease caused by HCV which comprises introducing into the tissue of a vertebrate the polynucleotide of Claim 1.

45                   10. A vaccine for inducing immune responses against HCV infection which comprises the polynucleotide of Claim 1 and a pharmaceutically acceptable carrier.

50                   11. A method for inducing anti-HCV immune responses in a primate which comprises introducing the polynucleotide of Claim 1 into the tissue of said primate and concurrently administering interleukin-12 parenterally.

- 23 -

12. A method of inducing an antigen presenting cell to stimulate cytotoxic and helper T-cell proliferation and effector functions including lymphokine secretion specific to HCV antigens which 5 comprises exposing cells of a vertebrate in vivo to the polynucleotide of Claim 1.
13. A method of treating a patient in need of such treatment comprising administering to the patient the polynucleotide of 10 Claim 1 in combination with interferon-alpha, Ribavirin, Zidovudine, or other pharmaceutically acceptable antiviral agents..
14. A pharmaceutical composition comprising the polynucleotide of Claim 1. 15
15. A method of inducing an immune response comprising administering the polynucleotide of Claim 1 to a patient, the administration of the polynucleotide antedating or coinciding or following administration to the patient of a subunit, recombinant, 20 recombinant live vector, inactivated, recombinant inactivated vector, or live attenuated HCV vaccine.
16. A method for inducing immune responses in a vertebrate against HCV epitopes which comprises introducing between 1 25 ng and 100 mg of the polynucleotide of Claim 2 into the tissue of the vertebrate.
17. A method for inducing immune responses against infection or disease caused by HCV which comprises introducing into 30 the tissue of a vertebrate the polynucleotide of Claim 2.
18. A vaccine for inducing immune responses against HCV infection which comprises the polynucleotide of Claim 2 and a pharmaceutically acceptable carrier.

- 24 -

19. A method for inducing anti-HCV immune responses in a primate which comprises introducing the polynucleotide of Claim 2 into the tissue of said primate and concurrently administering 5 interleukin 12 parenterally.

20. A method of inducing an antigen presenting cell to stimulate cytotoxic and helper T-cell proliferation an effector functions including lymphokine secretion specific to HCV antigens which 10 comprises exposing cells of a vertebrate in vivo to the polynucleotide of Claim 2.

21. A method of treating a patient in need of such treatment comprising administering to the patient the polynucleotide of 15 Claim 2 in combination with interferon-alpha, Ribavirin, Zidovudine, or other pharmaceutically acceptable antiviral agents..

22. A pharmaceutical composition comprising the 20 polynucleotide of Claim 2.

23. A method of inducing an immune response comprising administering the polynucleotide of Claim 2 to a patient, the administration of the polynucleotide antedating or coinciding or following administration to the patient of a subunit, recombinant, 25 recombinant live vector, inactivated, recombinant inactivated vector, or live attenuated HCV vaccine.

24. The vector of Claim 2 which is selected from V1Ra.HCV1CorePAb, Vtpa.HCV1CorePAb, VUb.HCV1CorePAb, 30 V1Ra.HCV1Core, Vtpa.HCV1Core and VUb.HCV1Core.

25. A pharmaceutical composition comprising the vector of Claim 21.

- 25 -

26. The DNA sequence of Claim 1 selected from the group consisting of a nucleotide sequence shown in Figure 5, Figure 9, Figure 10, Figure 11, Figure 12 and Figure 13.

1/22

1	10	20	30	40	50	60	70	80
	GATATTGGCT ATTGGCCATT	GCATACGGTTG	TATCCATATC	ATAATAATGTA	CATTTATATT	GGCTCATGTC	CAACATTACC	80
81	GCCATGTTGA	CATTGATTAT	TGACTAGTTA	TCAATTACGG	GGTCATTAGT	TCAAGCCC	TATATGGAGT	160
161	TCCGGCTTAC	ATAACTTACG	GTAAATGGCC	GGCCTGGCTG	ACGGCCAAC	GAACCCCGCC	CATTGACGTC	240
241	TATGTTCCA	TAGTAACGC	AATAGGGACT	TTCCATTGAC	GTCAATGGGT	GGAGTATTAA	CGGTAAACTG	320
321	AGTACATCAA	GTGTATCATA	TGCCAAGTAC	GCCCCCTATT	GACGTCAATG	ACGGTAATG	GCCCCCTGG	400
401	AGTACATGAC	CTTATGGGAC	TTCCCTACTT	GGCAGTACAT	CTACGTATT	GTCACTGCTA	TTACCATGGT	480
481	TGGCAGTACA	TCAATGGGG	TGGATAGGG	TTTGACTAAC	GGGGATTCC	AAGTCTAAC	CCCATTGACG	560
561	TTTGTGTTGG	CACCAAAATC	AACGGGACTT	TCCAAAATGT	CCTAAACAAT	CGGCCCAT	GACGAAATG	640
641	GTGTACGGTG	GGAGGTCTAT	ATAAGCAGAG	CTCGTTTACT	GAACCGTCAG	ATCGCCCTGA	ACGCTGT	720
721	GACCTCATA	GAAGACACCG	GGACCGATCC	AGCCTCGGG	GGCGGAAACG	GTGCAATTGA	ACGGGATTCC	800
801	GAGTGACGTA	AGTACCGCC	ATAGAGTCTA	TAGGCCACC	CCCTTGGCTT	CTTATGCTG	CTATACTGTT	880
881	GGTCTATACA	CCCCCGCTTC	CTCATGTTA	AGGTGATGGT	ATAGCTTAGC	CTATAGGTGT	GGTTTATTGA	960
961	CCACTCCCT	ATTGGGTGACG	ATACTTCCA	TACTAACTC	ATAACATGGC	TCTTGGCAC	AACTCTTT	1040
1041	TGCCAATACA	CTGTCCCTCA	GAGACTGACA	CGGACTCTGT	ATTTTACAG	GATGGGGTCT	CATTATTAT	1120
1121	ACATATACAA	CACCAACGCC	CCCAGTGGTC	GCAGTTTTA	TTAAACATAA	CCTGGGATCT	CCACCGAAT	1200
1201	TGTTCCGGAC	ATGGGCTCTT	CTCCGGTAGC	GGCGGAGCT	CTACATCGA	GCCCCTGCC	CATGCCCTCA	1280
1281	GTGCGCTGGC	AGCTCCTTGC	TCTTAACAGT	GGAGGCCAGA	CTTAGGCACA	GCACGATGCC	CACCAACC	1360
1361	ACAAGGGCGT	GGGGGTAGGG	TATGTGCTG	AAAATGAGCT	GGGGAGCGG	GCTTGGACCG	CTGACGCATT	1440
1441	AAGGAGCGG	CAGAAGAAGA	TGCAGGCCAGC	TGAGTTGTTG	TGTTCTGATA	AGAGTCAGAG	GTAACTCCCG	1520
1521	GTTAACGGTG	GAGGGCAGTG	TAGTCTGAGC	AGTACTCTT	GCTGGCGGC	GGGCCACAG	ACATAATAGC	1600
1601	ACAGACTGTT	CCTTTCCATG	GGTCTTTCT	GCAGTCACCG	TCCTTAGATC	TAGGTACAG	ATATCAGAA	1680
1680	GGGGCGCGA	TCTGCTGTGC	CTTCTAGTTG	CCAGCCCATCT	GTGTTTGCC	CCTCCCCGT	GCCTCCCTTG	1760
1761	GTGCCACTCC	CAGTGTCTCC	TCCTATAAA	ATGAGGAAAT	TGCATGCTAT	TGTCTGAGTA	GGTGTCAATT	1840
1841	GCTGGGGTGG	GGCAGCACAG	CAAGGGGGAG	GATTGGGAAG	ACAATAGCAG	GCATGCTGG	GATGGGGTGG	1920
1921	TACGGCGCA	GGGGCCTTA	TTAAGGCCG	AGCGGCCGC	CCCAGGTGCT	GAAGAATTGA	CCCGGTTCT	2000

FIG. 1A

2/22

2001	AAAGGCCGG	TTGCTGGCGT	TTTCCATAG	GCTCCGGCC	CCTGACGAGC	ATCACAAAAA	TCGACGCTCA	AGTCAGAGGT	2080		
2081	GGCGAAACCC	GACAGGACTA	TAAAGATACC	AGGGCTTCC	CCCTGGAAAGC	TCCCTCGTGC	GCTCTCTGT	TCCGACCCCTG	2160		
2161	CCGC TTACCG	GATACTGTC	GGCCTTTCTC	CCTTGGGA <sup>4</sup>	GGTGGCGCT	TTCTCAATGC	TCACGCTGTA	GGTATCTCAG	2240		
2241	TTGGGTGTA <sup>5</sup>	GTGCGTGCCT	CCAAGCTGGG	CTGTTGCAC	GAACCCCCCG	TTCAAGCCGA	CCGCTGGCC	TTATCCGGTA	2320		
2321	ACTATCGCT	TGAGTCCAAC	CGGTAAGAC	ACGACTTATC	GCCACTGGCA	GCAGCCACTG	GTAACAGGAT	TAGCAGAGCG	2400		
2401	AGGTATGTAG	GGGGTGCCT	AGAGTTCTG	AAGTGGTGGC	CTAACTACGG	CTACACTAGA	AGGACAGTAT	TTGGTATCTG	2480		
2481	CGCTCTGCTG	AAGCCAGTTA	CCTTCGGAAA	AAGAGTTGGT	AGCTCTTGTAT	CGGGCAAAACA	AACCACCGCT	GGTAGCGGGT	2560		
2561	GTTTTTTGT	TTGCAAGCAG	CAGATTACGC	GCAGAAAAAA	AGGATCTCAA	GAAGATCCTT	TGATCTTTTC	TAGTGTATCC	2640		
2641	CGTAATGCTC	TGCCAGTGT	ACAACCAATT	AACCAATTCT	GATTAGAAAAA	ACTCATCGAG	CATCAAATGA	AACTGCAATT	2720		
2721	TATTCCATTC	AGGATTATCA	ATACCATAAT	TTTGGAAAAAG	CGGGTTCTGT	AATGAAGGAG	AAAACCTACC	GAGGCAGTTC	2800		
2801	CATAGGATGG	CAAGATCCTG	GTATCGGTCT	GGGATTCCGA	CTCGTCCAAC	ATCAAAATCAA	CCTATTAAATT	TCCCCTCGTC	2880		
2881	AAAATAAGG	TTATCAAGTG	AGAAAATCACC	ATGAGTGCAG	ACTGAATTCG	GTGAGAATTGG	CAAAAGCTTA	TGCAATTCTT	2860		
2961	TCCAGACTTG	TTCAACAGGC	CAGCCATTAC	GCTGTCATC	AAAATCACTC	GCATCAACCA	AACCGTTATT	CATTCTGTAT	3040		
3041	TGCGCCTGAG	CGAGACGAAA	TACGGGATCG	CTGTTAAAG	GACAATTACA	AACAGGAATC	GAATGCAACC	GGGCAGGAA	3120		
3121	CACTGCCAGC	GCATCAACAA	TATTTTCAACC	TGAATCAGGA	TATTCTCTTA	ATACCTGGAA	TGCTGTTTTC	CGGGGGATCG	3200		
3201	CAGTGGTGA <sup>6</sup>	TAACCATGCA	TCATCAGGAG	TACGGATAAA	ATGGTTGATG	GTGGAAAGAG	GCATAAAATTC	CGTCAGCCAG	3280		
3281	TTTAGTCTGA	CCATCTCATC	TGTAACATCA	TTGGCAACGC	TACCTTTGCC	ATGTTTCAGA	AACAACCTTG	GGGCATCGGG	3360		
3361	CTTCCCACATAC	AATCGATAGA	TTGTCGCAACC	TGATTGCCCG	ACATTATCGC	GAGGCCATT	ATACCCATAT	AAATCAGGCAT	3440		
3441	CCATGTTGGA	ATTTAATCGC	GGGCCTCGAGC	AAGACGTTTC	CGGTTGAATA	TGGCTCATAA	CACCCCTGT	ATTACTGTTT	3520		
3521	ATGTAAGCAG	ACAGTTTTAT	TGTTCATGAT	GATATATTTT	TATCTGTGC	AATGTAACAT	CAGAGATT	GAGACACAAC	3600		
3601	GTGGCTTCC	10	10	20	20	30	40	50	60	70	80

FIG. 1B

3/22

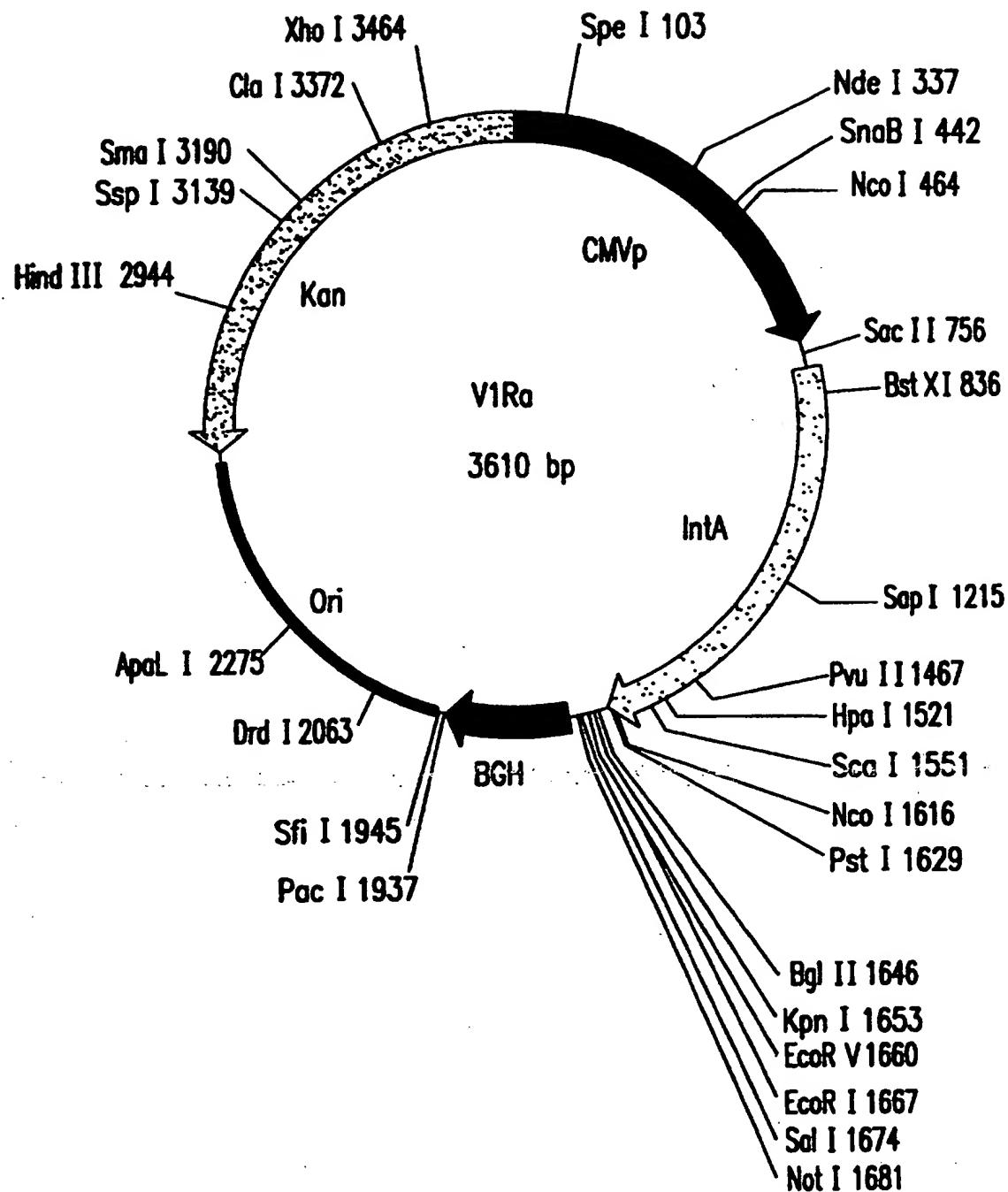


FIG.2

4/22

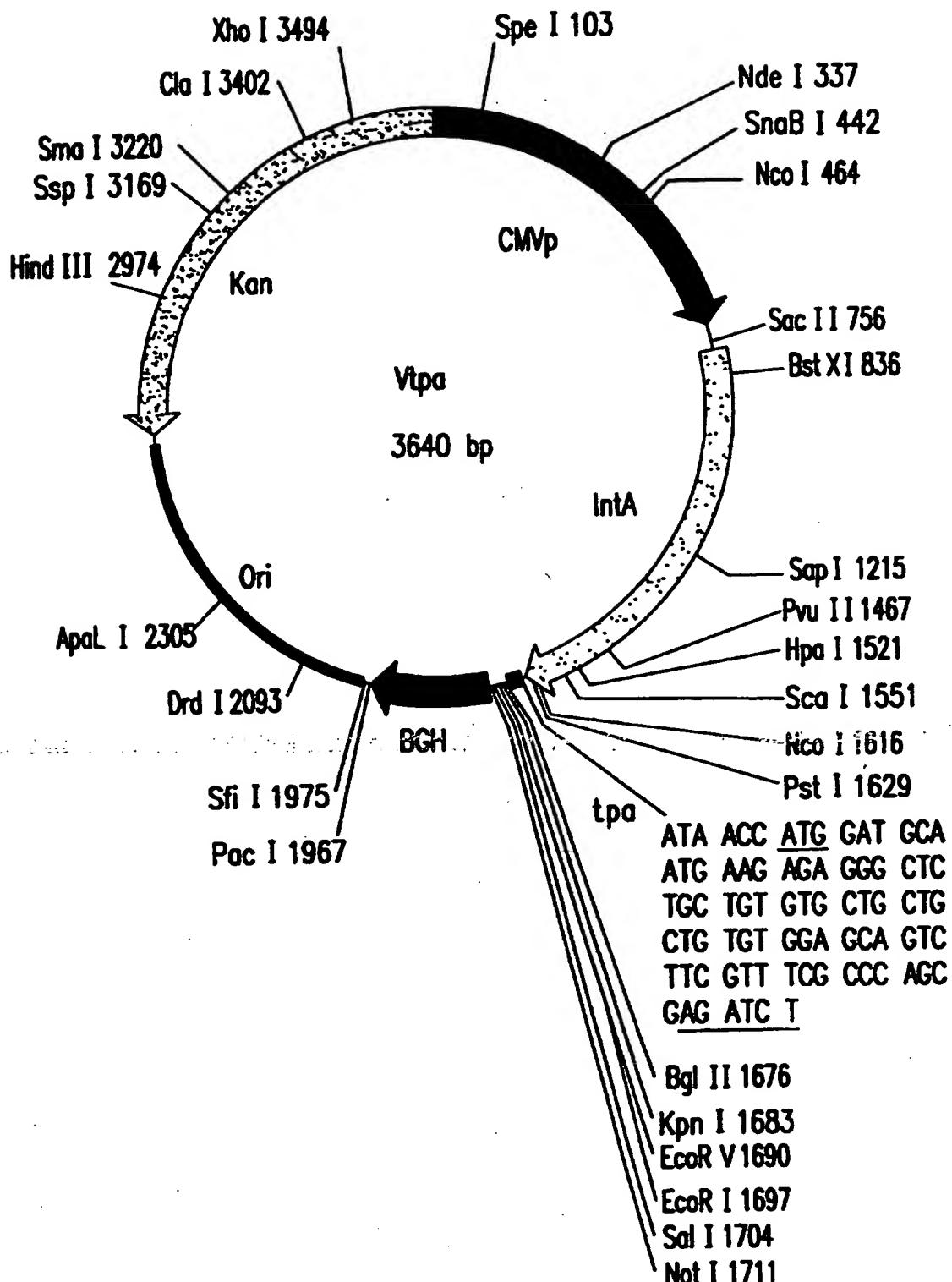
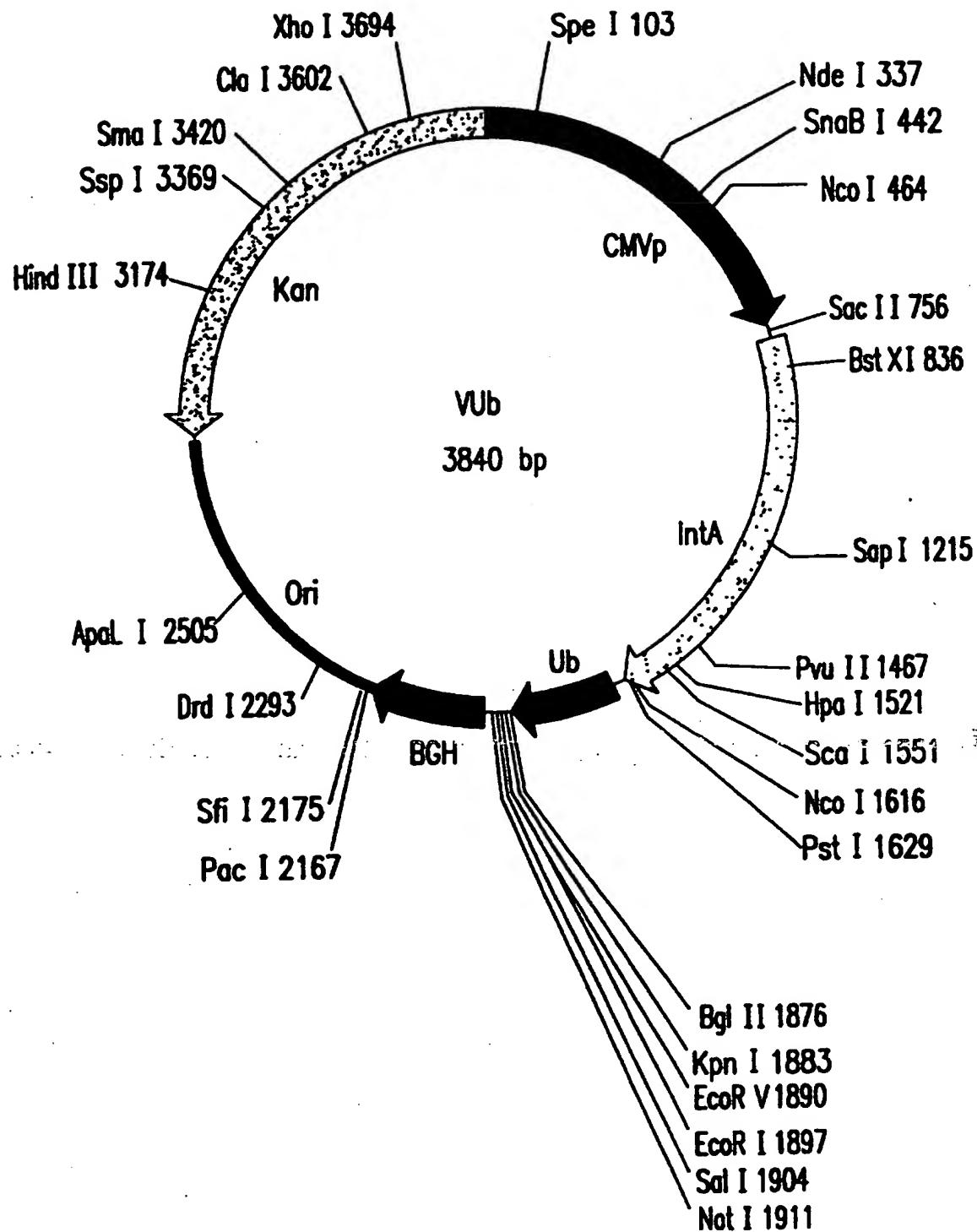


FIG.3

5/22



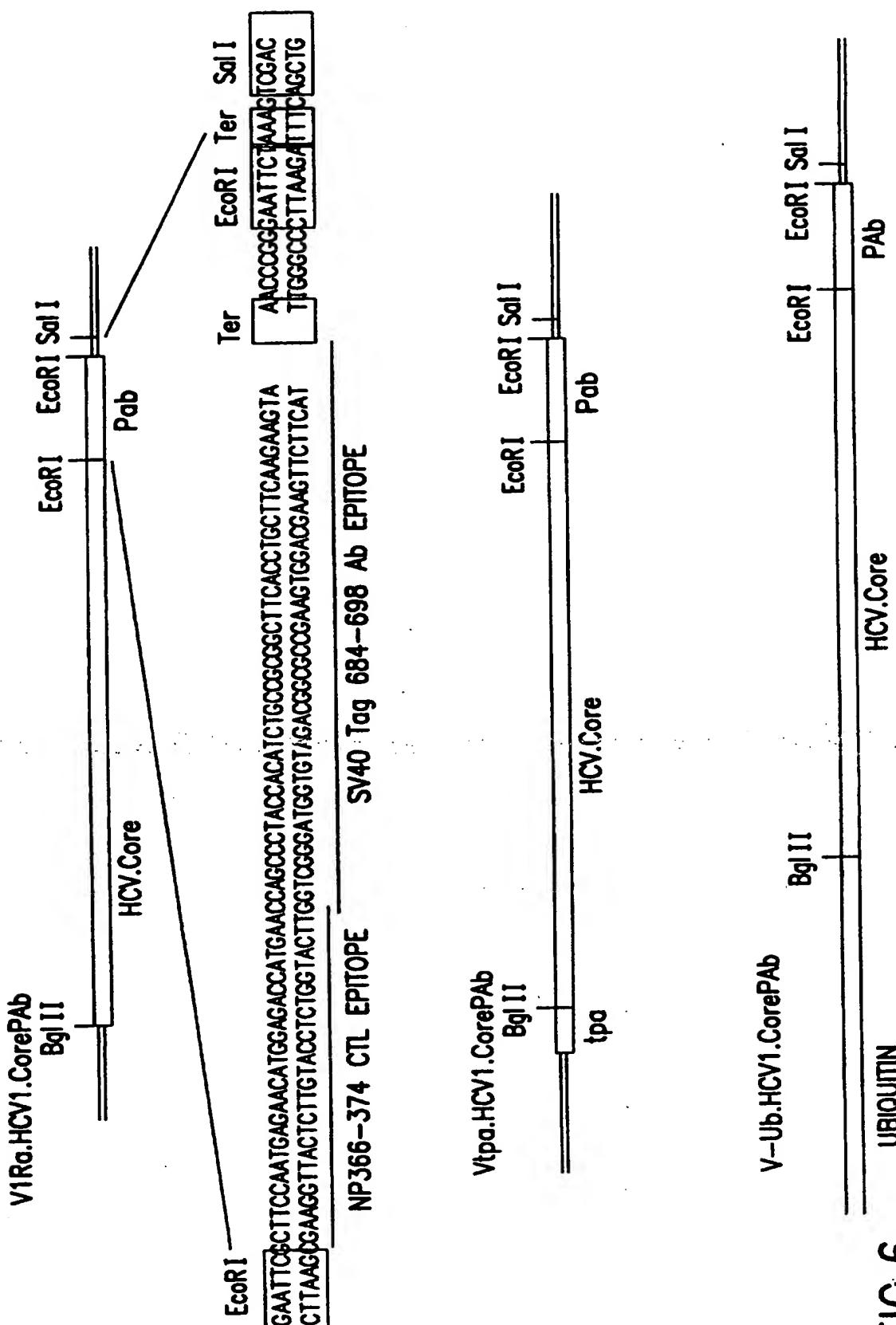
**FIG.4**

6/22

1/1	ATG	AGC	ACC	AAC	CCC	AAg	CCC	CAG	AGg	AAG	ACC	AAc	AAg	AGg	AGg	CCCAG	31/11				
Met	ser	thr	asn	pro	lys	pro	ala	arg	lys	thr	lys	arg	asn	thr	asn	arg	arg	pro	gln		
61/21																			91/31		
	GAT	GTg	AAG	TTC	CCT	GGg	GGa	GGC	CAG	ATT	GTg	GGa	GGg	GTc	TAC	CTG	CTG	CTG	AGG		
	asp	val	lys	phe	pro	gly	gly	gly	gln	11e	val	gly	gly	val	gly	val	lys	leu	leu		
	121/41																		151/51		
	GGC	CCC	AGG	CTG	GGg	GTG	agg	GGC	ACC	agg	AAG	ACC	TCT	GAG	AGG	TCC	CAG	CCC	AGg		
	gly	pro	arg	leu	gly	val	arg	ala	thr	arg	lys	thr	ser	glu	arg	ser	gln	pro	arg		
	181/61																		211/71		
	AGG	agg	CAG	CCC	ATC	CCC	AAG	GGC	agg	GGG	CCT	GAG	GGC	cGC	TCC	TGG	GGC	CAG	CCT	GGC	
	arg	arg	gln	pro	ile	pro	lys	ala	arg	arg	arg	pro	glu	gly	arg	ser	trp	ala	gln	pro	gly
	241/81																		271/81		
	TAC	CCC	TGG	CCC	CTg	TAT	GGC	AAT	GAA	GGC	TTt	GGC	TGG	GCT	GGC	TGG	CTG	CTG	TCC	CCC	
	try	pro	trp	pro	leu	tyr	gly	asn	glu	gly	trp	ala	gly	trp	ala	gly	trp	leu	leu	ser	pro
	301/101																		331/111		
	agg	GGC	TCC	agg	CCC	tcc	TGG	GGC	CCC	ACA	GAC	CCC	agg	agg	AGG	TCC	agg	AAc	cTG	GGC	
	arg	gly	ser	arg	pro	ser	trp	gly	pro	thr	asp	pro	arg	arg	arg	ser	arg	asn	leu	gly	
	361/121																		391/131		
	AAg	GTg	ATT	GAC	ACC	CTg	ACC	TGt	GGC	TTt	GCT	GAC	CTg	ATG	GGC	TAC	ATC	CCC	CTg	GTg	
	lys	val	ile	asp	thr	leu	thr	cys	gly	phe	ala	asp	leu	met	gly	tyr	ile	pro	leu	val	
	421/141																		451/151		
	GGg	GCT	CCT	GTg	GGg	GGg	GTg	GCT	AGG	GCT	CTG	GCT	CAT	GGg	GTg	AGG	GTg	CTG	GAG	GAT	
	gly	ala	pro	val	gly	gly	val	ala	arg	ala	leu	ala	his	gly	val	arg	val	leu	glu	asp	
	481/161																		511/171		
	GGG	GTg	AAC	TAT	GCT	ACT	GGC	AAC	CTG	CCT	GGc	TGC	TCC	TTG	TCC	ATC	TTC	CTg	CTG	GGC	
	gly	val	asn	tyr	ala	thr	gly	asn	leu	pro	gly	cys	ser	phe	ser	ile	phe	leu	leu	ala	
	541/181																		571/191		
	CTG	CTC	TCC	TGC	CTG	ACa	GTg	CCT	GCT	TCT	GGC										
	leu	leu	ser	cys	leu	thr	val	pro	ala	ser	ala										

FIG. 5

7/22



6  
FIG

8/22

1/1 ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAC CGT AAC ACC CGC CGC CCA CAG  
 Met ser thr asn pro lys pro glu arg lys thr lys arg asn thr asn arg arg pro glu  
 61/21 GAC GTC AAg TTC CCG GGC GGT CAG ATC GTT GGT GGA GTT TAC TTC TTG CCG CGC AGG  
 asp val lys phe pro gly gly gly glu lys val gly gly val lys arg arg pro glu  
 121/41 GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT TCC GAG CGG TCG CAA CCT CGT GGA  
 gly pro arg leu gly val arg ala thr arg lys thr lys arg ser glu arg pro arg gly  
 181/61 AGG CGa CAG CCT ATC CCC AAG GCT CGc CGG CCC GAG GGC AGG TCC TGG GCT CAG CCC GGG  
 arg arg glu pro ile pro lys ala arg arg pro glu gly arg ser trp ala glu pro gly  
 241/81 TAC CCT TGG CCC CTC TAT GAg GGC Ttc GGG TGG GCA GGA TGG CTC CTG TCC CCC  
 tyr pro trp pro leu tyr gly asn glu gly trp ala gly trp ala glu trp ala glu pro gly  
 301/101 CGC GGC TCT CGg CCT agT TGG GGC CCC ACT GAC CCC CGG CGt AGG TCG CGC AAT TTG GGT  
 arg gly ser arg pro ser trp gly pro thr asp pro arg arg ser arg asn leu gly  
 361/121 AAG GTC ATC GAT ACC CTC ACG TGC GGC TTC GGC GAC CTC ATG GGG TAC ATC CCG CTC GTC  
 lys val ile asp thr leu thr cys gly phe ala asp leu met gly tyr ile pro leu val  
 421/141 451/151 GGC CCC GTC GGG GGC GTC GGC AGg GGC CTG GCG CAT GGC GTC AGG GtT CTG GAG GAC  
 gly ala pro val gly gly val ala arg ala leu ala his gly val arg val leu glu asp  
 481/161 511/171 GGG gtg AAC TAT GCA ACA GGG AAt tTg CCC GGT TGC TCT TTC TCT ATC TTC CTC CTG GCT  
 glu val asn tyr ala thr gly asn leu pro gly cys ser phe ser ile phe leu leu ala  
 541/181 571/191 CTg CTg TCC TGC CTG ACC GTC CCA GCT TCT GCT  
 leu leu ser cys leu thr val pro ala ser ala

**FIG. 7**

9/22

TABLE 3  
CODON UTILIZATION IN HUMAN PROTEIN-CODING SEQUENCES

	a	b	c	d	e	f		a	b	c	d	e	f							
F	UUU	68	0.35	193	4.5		Y	UAU	72	0.47	153	3.6								
	UUC	125	0.65					UAC	81	0.53										
L	UUA	20	0.05	445	10.4		H	CAU	44	0.42	105	2.5								
	UUG	42	0.09					CAC	61	0.58										
	CUU	50	0.11				Q	CAA	50	0.26	192	4.5								
	CUC	99	0.22					CAG	142	0.74										
	CUA	30	0.07				N	AAU	51	0.34	148	3.5								
	CUG	204	0.46					AAC	97	0.66										
I	AUU	28	0.23	123	2.9				K	AAA	137	0.45	303	7.0						
	AUC	79	0.64						AAG	166	0.55									
	AUA	16	0.13				D	GAU	79	0.38	209	4.9								
M	AUG	77	1.00	77	1.8				GAC	130	0.62									
V	GUU	35	0.13	266	6.2		E	GAA	125	0.40	311	7.3								
	GUC	72	0.27					GAG	186	0.60										
	GUA	25	0.09				C	UGU	44	0.30	147	3.4								
	GUG	134	0.50					UGC	103	0.70										
S	UCU	59	0.17	349	8.1		W	UGG	56	1.00	56	1.3								
	UCC	91	0.26						CGV	19	0.09	215	5.0							
	UCA	37	0.11				R	CGC	40	0.19										
	UCC	25	0.07					CGA	22	0.10										
	AGU	37	0.11					CGG	33	0.15										
	AGC	100	0.29					AGA	51	0.24										
P	CCU	51	0.24	212	4.9				AGG	50	0.23									
	CCC	86	0.41				G	GGU	36	0.15	245	5.7								
	CCA	51	0.24					GGC	108	0.44										
	CCG	24	0.11					GGA	42	0.17										
T	ACU	47	0.20	238	5.6				GGG	59	0.24									
	ACC	113	0.47				A	TOTAL 4285 RESIDUES EXCLUDING N-TERMINAL METHIONINE RESIDUES												
	ACA	50	0.21																	
	ACG	28	0.12																	
A	GCU	91	0.31	298	7.0															
	GCC	119	0.40																	
	GCA	51	0.17																	
	GCC	37	0.12																	

FIG.8

10/22

1/1 31/11 atg TAT GAG GTG aGg AAt GTC TCT GGC GTC TAC CAT GTg ACC AAt GAC TGC TCC AAC TCC  
 M Y E V R N V S G V Y H V T N D C S N S  
 61/21 tGc ATT GTC TAT GAG GCT GAC ATG ATC ATG CAC ACC CCT GGC TGT GTg CCa TGT GTg  
 C I V Y E A D M I M H T P G C V P C V  
 121/41 aGG GAG GGC AAC TCC TCC aGg TGC TGG GTg GCC CTg ACC CCC ACC CTg GCT GCC AGG AAC  
 R E G N S S R C W V A L T P T L A A R N  
 181/61 tCC tCC ATC CCC ACC ACC ATC aGg aGg CAT GTg GAC CTG CTg GTg GGC GCT GCT GCC  
 S S I P T T T I R R H V D L L V G A A A  
 241/81 CTg TGC Tct Gcc ATG Tat GTG GGC GAC CTg TGT GGC TCT GTC TCC CTg GTg TCC CAG gTG  
 L C S A M Y V G D L C G S V F L V S Q L  
 301/101 TTC ACC TTC TCC CCC aGg aGG TAT GAG ACT GTg CAG GAC TGC AAC TGC TCC CTg TAC CCT  
 F T S P R R Y E T V Q D C N C S L Y P  
 361/121 GGC CAT GTC Tct GGC CAC aGg ATG GCC TGG GAC ATG ATG AAC TGG TCC CCC ACC ACT  
 G H V S G H R M A W D M M M N W S P T T  
 421/141 GGC CTg GTG GTC TCC CAG CTg aGG ATT CCC CAG GCT GTg GAC ATG GTG TGT GGG  
 A L V S Q L L R I P Q A V V D M V V G  
 481/161 GGC CAC TGG GGC GTg CTG GCT GGC CTg GCC TAC TAC TCC ATG GTG GGC AAC TGG GCC AAG  
 A H W G V L A G L A Y S M V G N W A K  
 541/181 GTg CTG ATT GTG ATG CTg CTg TTT GCT GGC GTg GAt GGC taA  
 V L I V M L L F A G V D G \* FIG. 9  
 571/191

11/22

1/1 31/111  
 atg ACC ACC TAT GTC TCT GTG GGC CAT GCC tcc CAG ACC ACC aGG aGG GTg GCC TCC TCC  
 M T T Y V S V G H A S Q T T R R V A S F  
 61/21 51/31

TTC tcc CCT GGC TCT GCC CAG AAG ATC CAG CTg GTg AAC ACC AAT GGC tcc TGG CAC ATC  
 F S P G S A Q K I Q L V N T N G S W H I

121/41 151/51  
 AAC AGG ACT GCC CTG AAT TGC AAT GAG TCC ATC AAC ATC GGC TTC TTT GCT GCC CTG TCC  
 N R T A L N C N E S I N T G F F A A L F

181/61 211/71  
 TAT GTg AAG AAG TTC AAC TCC TCT GGC TGC TCT GAG aGG ATG GCC tct TGC aGG CCC ATT  
 Y V K F N S G C S E R M A S C R P I

241/81 271/91  
 GAC AGG TTT GCC CAg GGC TGG GGC CCC ATC ACC CAT GCT GAG TCC aGG tcc TCT GAC CAG  
 D R F A Q G W G P I T H A E S R S D Q

301/101 331/111  
 AGG CCA TAC TGC TGG CAC TAT GCC CCC CAG CCA TGT GGC ATT GTG CCT GCC CTG CAT GTC  
 R P Y C W H Y A P Q P C G I V P A L H V

361/121 391/131  
 Tgt GGC CCT GTC TAC TGC TTC ACC CCA tcc CCT GTg GTg GGC ACg ACT GAC aGg TTt  
 C G P V Y C F T P S P V V G T T D R F

421/141 451/151  
 GGC GTg CCC ACC TAC AAC TGG GGC GAC AAT GAG ACT GAT GTG CTg CTg AAC AAC ACC  
 G V P T Y N W G D N E T D V L L N N T

481/161 511/171  
 aGG CCC CCC CAG GGC AAC TGG TTT GGC TGC ACC TGG ATG AAC tcc ACT GGC TTC ACC AAG  
 R P Q G N W F G C T W M N S T G F T K

FIG. 10A

12/22

541/181	ACC Tgt GGC CCC Cca Tgc AAC ATT GGC GGC GCT GGC AAC AAC ACC CTG ACC TGC CCC	571/191	GGC GCT GGC Tct GGC TCC ACC AAG TGT GGC Tct GGC CCA TGC CCC
T C G G G G	P P C N I G G A G N N T L T C P	601/201	T G H P E A T Y T K C G S G P W
T D C F R K	631/211	T	661/221
L T P R C M V D Y P Y R L W H Y P C T F	691/231	CTG ACC CCC AGG TGC ATG GTg GAC TAC CCA TAC AGg CTg TGG CAC TAC CCA TGC ACC TTC	
N F T I F K I R M Y V D Y R L W H Y P C T F	751/251	CTG ACC ACC ATC TTC AAG ATC AGG ATG TAT GTG GGC GGC GTG GAG CAC AGG CTg AAT GCT	
A C N W T R G E R C N I E D R D R S E L	811/271	AAC TTC ACC ATC TTC AAG ATC AGG ATG TAT GTG GGC GAG TCT GAG AGG GAC AGG GAC AGG GAG GCT	
A C N W T R G E R C N I E D R D R S E L	871/291	GCC TGC AAC TGG ACC aGg GGC GAG aGg TGC AAC ATT GAG GAC AGG GAC AGG GAC AGG GAG GCT	
S P L L S T T E W Q I L P C S F T T L	931/311	tcc CCC CTG CTg CTG TCC ACC Act GAG TGG CAG ATC CTg CCa TGC TCC TTC ACC ACC CTG	
Cct GCC CTG TCC ACT GGC CTG ATC CAT CTg CAT CAG AAC ATT GTG GAT GTG CAG TAC CTG	901/301	S A L S T G L I H L H Q N I V D V Q Y L	
P A L S T G L I H L H Q N I V D V Q Y L	961/321	TAt GGC GTg GGC Tct Gct GTg GTC TCC ATT GTG ATC AAg TGG GAG TAT GTg CTG CTG CTG	
Y G V G S A V V S I V I K W E Y V L L L	1021/341	TTT CTg CTg CTG GCT GAT GCC taa	
F L L L A D A *			

## FIG. 10B

13/22

1/1	TAT	GAG	GTG	aGg	AAt	GTC	TCT	GGC	GTC	TAC	CAT	GTg	ACC	AAt	GAC	TGC	TCC	AAC	TCC	31/11
M	Y	E	V	R	N	V	S	G	V	Y	H	V	T	N	D	C	S	N	S	61/21
tGc	ATT	GTC	TAT	GAG	GCT	GAC	ATG	ATC	ATG	CAC	ACC	CCT	GGC	TGt	GTg	CCa	TGt	GTg	91/31	
C	I	V	Y	E	A	A	D	M	I	M	H	T	P	G	C	V	P	C	V	121/41
aGG	GAG	GGC	AAC	TCC	TCC	aGg	TGC	TGG	GTg	GGC	CTG	ACC	CCC	ACC	CTG	GCT	GCC	AGG	AAC	151/51
R	E	G	N	S	S	R	C	W	V	A	L	T	P	T	L	A	A	R	N	181/61
tCC	tCC	ATC	CCC	ACC	ACC	ATC	aGg	aGg	CAT	GTg	GAC	CTG	CTG	GTg	GGC	GCT	GCT	GCC	211/71	
S	S	I	P	T	T	T	I	R	R	H	V	D	L	L	V	G	A	A	A	241/81
CTg	TGC	TCT	GCC	ATG	TAT	GTG	GGC	GAC	CTg	TGT	GGC	TCT	GTC	TTC	CTG	GTg	TCC	CAG	CTG	271/91
L	C	S	A	M	Y	V	G	D	L	C	G	S	V	F	L	V	S	Q	L	301/101
TTC	ACC	TTC	TCC	CCC	aGg	agg	TAT	GAG	ACT	GTg	CAG	GAC	TGC	AAC	TGC	TCC	CTg	TAC	CCT	331/111
F	T	S	P	R	R	Y	E	T	V	Q	D	C	N	C	S	L	Y	P	361/121	
GGC	CAT	GTC	TCT	GGC	CAC	aGg	ATG	GCC	TGG	GAC	ATG	ATG	AAC	TGG	TCC	CCC	ACC	ACT	451/151	
G	H	V	S	G	H	R	M	A	W	D	M	M	N	W	S	P	T	T	421/141	
GGC	CTg	GTG	GTC	TCC	CAG	CTg	aGg	ATT	CCC	CAG	GCT	GTg	GAC	ATG	GTG	GTG	GGC	511/171		
A	L	V	S	Q	V	L	R	R	P	Q	A	V	D	M	V	V	G	481/161		
GCC	CAC	TGG	GGC	GTg	CTG	GCT	GGC	CTg	GGC	TAC	TAC	TCC	ATG	GTG	GGC	AAC	TGG	GCC	AAG	511/171
A	H	W	G	V	L	A	G	L	A	Y	S	M	V	G	N	W	A	K	481/161	

FIG. 11A

14/22

541/181 GTg cTg ATT GTG ATG CTg CTg TTT Gct GGC GTg GAt GGC ACC ACC TAt GTC TCT GTG GGC  
 V L I V M L F A G V D G T T Y V S V G  
 601/201 CAT GCC tCC CAG ACC ACC aGG aGG GTg GCC TCC TTC TCC tCC CCT GGC TCT GCC CAG AAG  
 H A S Q T T R R V A S F F S P G S A Q K  
 661/221 ATC CAg CTg GTg AAC ACC AAt GGC tCC TGG CAC ATC AAC AGG ACT GCC CTG AAt TGC AAt  
 I Q L V N T N G S W H I N R T A L N C N  
 721/241 GAG TCC ATC AAC ACT GGC TTC TTT Gct GCC CTG TTC TAt GTg AAG AAG TTC AAC TCC TCT  
 E S I N T G F F A A L F Y V K K F N S S  
 781/261 GGC TGC TCT GAG aGg ATG GCC tct TGC aGg CCC ATT GAC AGG TTt GCC CAg GGC TGG GGC  
 G C S E R M A S C R P I D R F A Q G W G  
 841/281 CCC ATC ACC CAT GCT GAG TCC aGG tCC Tct GAC CAG AGG CCA TAC TGC TGG CAC TAT GCC  
 P I T H A E S R S S D Q R P Y C W H Y A  
 901/301 CCC CAg CCa TGT GGC ATT GTG CCT GCC CTG CAT GTC TGT GGC CCT GTC TAC TGC TTt ACC  
 P Q C G I V P A L H V C G P V Y C F T  
 961/321 CCa tCC CCT GTg GTg GGC ACC ACT GAC aGg TTt GGC GTg CCC Acc TAC AAC TGG GGC  
 P S P V V G T D R F G V P T Y N W G  
 1021/341 GAC AAT GAG ACT GAT GTG CTg CTg AAG AAC ACC aGG CCC CCC CAg GGC AAC TGG TTe  
 D N E T D V L L N N T R P P Q G N W F

FIG. 11B

1081/361 1111/37  
 GGC TGC ACC TGG ATG AAC tcc ACT GGC ACC AAG ACC TGT GGC GGC CCC CCC TGC AAC  
 G C T W M N S T G F T K T C G G P P C N  
 1141/381 1171/391  
 ATT GGC GGC GCT GGC AAC ACC CTG ACC TGC CCC ACT GAC TGC TTC aGG AAG CAT CCT  
 I G G A G N N T L T C P T D C F R K H P  
 1201/401 1231/411  
 GAG GCC ACC TAC ACC AAG TGT GGC TCT GGC CCA TGG CTG ACC CCC AGG TGC ATG GTg GAC  
 E A T Y T K C G S G P W L T P R C M V D  
 1261/421 1291/431  
 TAC CCA TAC AGg CTg TGG CAC TAC CCA TGC ACC TTC AAC TTC ACC ATC TTC AAG ATC AGG  
 Y P Y R L W H Y P C T F N F T I F K I R  
 1321/441 1351/451  
 ATG TAT GTG GGC GGC GTG GAG CAC AGG CTg AAT GCT GCC TGC AAC TGG ACC aGG GGC GAG  
 M Y V G G V E H R L N A A C N W T R G E  
 1381/461 1411/471  
 aGG TGC AAC ATg GAG CAC AGG GAC AGG TCT GAG CTg tcc CCC CTG CTg CTG TCC ACC ACT  
 R C N I E D R D R S E L S P L L S T T  
 1441/481 1471/491  
 GAG TGG CAG ATC CTg CCA TGC TCC ACC ACC CTg CCT GCC CTG TCC ACT GGC CTG ATC  
 E W Q I L P C S F T T L P A L S T G L I  
 1501/501 1531/511  
 CAT CTg CAT CAG AAC ATT GTG GAT GTG CAG TAC CTG TAT GGC GTg GGC TCT GCT GTg GTc  
 H L H Q N I V D V Q Y L Y G V G S A V V  
 1561/521 1591/531  
 TCC ATT GTG ATC AAG TGG GAG TAT GTg CTg CTg GCT GAT GCC taa  
 S I V I K W E Y V L L F L L A D A \*

**FIG. 11C**

16/22

1/1	Tct	GGc	TCC	TGG	CTg	AGG	GAT	GTC	TGG	ATC	TGC	Act	GTG	CTG	ACT	GAC	TCC		
atg	M	S	G	S	W	L	R	D	V	W	D	I	C	T	V	L	T		
61/21																			
AAG	ACC	TGG	CTg	CAT	TCC	AAG	CTg	CTG	CCC	AGG	CTG	CCT	GGc	GAC	CCa	TTC	TCC	TGc	
K	T	W	L	H	S	K	L	L	P	R	L	P	G	D	P	F	F	C	
121/41																			
CAg	aGg	GGc	TAC	AGG	GGc	GTC	TGG	AGG	GGc	GAT	GGc	GTg	ATG	CAG	ACC	ACC	TGC	CCa	TGT
Q	R	G	Y	R	G	V	W	R	G	D	G	V	M	Q	T	T	C	P	C
181/61																			
GGc	GGc	CAG	ATC	ACT	GGc	CAT	GTg	AAG	AAT	GGc	TCC	ATG	AGG	ATT	GTg	GGc	CCC	AAG	ACC
G	A	Q	I	T	G	H	V	K	N	G	S	M	R	I	V	G	P	K	T
241/81																			
TGC	tCC	AAC	ACC	TGG	CAT	GGc	ACC	TTC	CCC	ATC	AAT	GCC	TAC	ACC	Act	GGc	CCa	TGC	ACC
C	S	N	T	W	H	G	T	F	P	I	N	A	Y	T	T	G	P	C	T
301/101																			
CCa	TCC	CCT	GGC	CCC	AAC	TAC	TCC	AGG	GGc	CTG	TGG	aG	GTG	GCT	GCT	GAG	GAT	TAT	GTG
P	S	P	A	P	N	Y	S	R	A	L	W	R	V	A	E	Y	V	V	V
361/121																			
GAG	GTg	Acc	aGg	GTG	GGc	GAC	TTC	CAC	TAT	GTG	Act	GGc	ATG	ACC	ACT	GAC	AAt	GTg	AG
E	V	T	R	V	G	D	F	H	Y	V	T	G	M	T	T	D	N	V	K
421/141																			
TGC	CCa	TGC	CAG	GTg	CCT	GGC	CCT	GAG	TTC	TTC	Act	GAG	GTG	GAT	GGc	GTG	aGG	CTG	CAC
C	P	C	Q	V	P	A	P	E	F	F	T	E	V	D	G	V	R	L	H
481/161																			
AGG	TAT	GCC	CCT	GGC	TGC	AAG	CCC	CTg	CTg	GAT	GAG	GTg	ACC	TTC	CAG	GTg	GGC	CTg	
R	Y	A	P	A	C	K	K	P	L	R	D	E	V	T	F	Q	V	G	

FIG. 12A

17/22

541/181 AAC CAG TTC CCT GTg GGC TCC CAG CTg CCa TGT GAG CCT GAG CCT GAT GTg ACT GTg CTg N Q F P V G S Q L P C E P D V T V L 601/201 ACC TCC ATG CTg ACT GAG CCA TCC CAC A:c ACT GCT GAG ACT GCC AAG aGg AGG CTG GCC T S M L T E P S H I T A E T A K R R L A. 661/221 AGg GGC TCC CCT CCA TCC CTg GCC tcc TCC TGCC tcc CAG CTG TCT GCT CCa TCC CTG R G S P P S L A S S A S Q L S A P S L 691/231 721/241 AAG GCC ACC TGC ACC ACC aGg CAT GAC TCC CCT GAT GCT GAC CTg ATT GAG GCC AAC CTg K A T C T R H D S P D A D L I E A N L 751/251 781/261 CTG TGG aGG CAG GAG ATG GGC GGC AAC ATC ACC aGG GTG GAG Tct GAG AAC AAG GTg GTg L W R Q E M G N I T R V E S E N K V V 841/281 ATC CTg GAC TCC TTT GAg CCC CTg aGG Gct GAG GAG GAT GAG AGG GAG GTC Tct GTG Gct I L D S F E P L R A E D E R E V S V A 901/301 Gct GAG ATC CTg aGG AAg tcc AGG AAG TTC CCC CCT Gcc CTG CCC ATc TGG GCG aGg CCa A E I L R K S R K F P P A L P I W A R P 961/321 tcc TAC AAC CCa CCC CTg CTg GAG TCC TGG AAG GAC CCT GAC TAT GTg CCC CCT GTG GTg S Y N P P L L E S V D P D Y V P P V V 1021/381 Cat GGC TGC CCC CTG CCC ACC ATG GCC CCa CCC ATc CCC CCa CCC aGg AGG AAG AGG H G C P L P P T M A P P I P R R K R

571/191 871/291 931/311 991/331 1051/371

FIG. 12B

18/22

1081/361	ACT GTg GTg CTG ACT GAg TCC Act GTC Tcc TCT GCC cTg GCT GAG CTg GCC ACC AAG ACC	1111/371
T V V L T E S T V S T V S A L A E L A T K T		
1141/381	TTC GGC tCC TCT GGC TCC TCT Gct GTg GAC tct GGC ACT GCC ACG GCC CCC CCT GAC CAG	1171/391
F G S S G S A V D S G T A T A P P D Q		
1201/401	Cca TCT GAT GAT GGC GAC AGg GGC Tct GAT GAT GAG TCC TAC TCC ATG CCC CCC CTg	1231/411
P S D D G D R G S D D E S Y S S M P P L		
1261/421	GAG GGC GAG CCT GGC GAC GAC CTg tct GAT GGC TCC TGG TCC ACT GTC tct GAG GAG	1291/431
E G E P G D P D S D G S W S T V S E E		
1321/441	Gcc tct GAG GAT GTg Gcc TGC TGC Tcc taa	*
A S E D V A C C S		

FIG. 12C

19/22

31/11  
 ATG TCC TAC ACC TGG ACT GCC CTg ATC ACC CCa TGT GCT GAG GAG tcc AAG CTG  
 M S Y T W T G A L I T P C A A E E S K L  
 91/31

61/21  
 CCC ATC AAC CCC CTG tcc AAC TCC CTG CTG aGG CAT CAC AAC ATG GTC TAT GCC ACC ACC  
 P I N P L S N S L L R H H N M V Y A T T  
 151/51

121/41  
 TCC aGg tct GCT GCC CTg aGG CAG AAG GTg ACC TTT GAC AGg CTG CAT GTg CCT GAT  
 S R S A G L R Q K V T F D R L H V P C

181/61  
 GAC CAC TAC aGG GAT GTG CTg AAG GAG ATG AAG GCC AAG GCC TCC ACT GTg AAG GCG AAG  
 D H Y R D V L K E M K A K S T V K A K  
 271/91

241/81  
 CTg CTg TCT GTg GAG GAG GCC TGC AAG CTG ACC CCT CCC CAC Tct GCC AGg TCC AAg TTT  
 S V E A C K L T P H S A R S K F

301/101  
 GGC TAT GGC GCC AAG GAT GTg aGG AAC CTg TCC tcc AAG Gct GTg AAC CAC ATC CAC TCT  
 G Y G A K D V R N L S S K A V N H I H S  
 391/131

361/121  
 GTC TGG AAG GAC CTG CTG GAG GAC ACT GAG ACC CCC ATT GAC ACC ACC ATG GCC AAg  
 V W K D L L E D T E T P I D T T I M A K

421/141  
 AAT GAG GTC TTC TGT GTg CAg CCT GAG AAG GGC GGC aGG AAG CCT GCC aGG CTg ATT GTC  
 N E V F C V Q P E K G G R K P A R L I V  
 511/171

481/161  
 TTC CCT GAG CTg GCC GTg aGG GTG TGT GAG AAg ATG GCC CTg TAT GAT GTG GTC TCC ACC  
 F P E L G V R C E K M A L Y D V V S T

FIG. 13A

20/22

541/181	CTg CCC CAG Gct GTG ATG GGC TCC TCC TAT GGC TTC CAG TAC TCC CCT GGC CAG aGG GTg	571/191	CTg CCC CAG Gct GTG ATG GGC TCC TCC TAT GGC TTC CAG TAC TCC CCT GGC CAG aGG GTg	
L P Q A V M G S S Y G F Q Y S P G Q R V	601/201	GAG TTC CTG GTG AAT GCC TGG AAg TCC AAG AAC CCC ATG GGC TTT GCC TAC TGC ACC	631/211	GAG TTC CTG GTG AAT GCC TGG AAg TCC AAG AAC CCC ATG GGC TTT GCC TAC TGC ACC
E F L V N A W K S K N P M F A Y C T	661/221	aGg TGC TTT GAC TCC ACT GTg ACT GAG tCT GAC ATC aGg GTg GAG TCC ATC TAC CAg	691/231	aGg TGC TTT GAC TCC ACT GTg ACT GAG tCT GAC ATC aGg GTg GAG TCC ATC TAC CAg
R C F D S T V T E S D I R V E S I Y Q	721/241	TGC TGT GAC CTG GCT CCT GAG GCC Agg CAG GTg ATC AGG TCC CTg ACT GAG aGG CTg TAC	751/251	TGC TGT GAC CTG GCT CCT GAG GCC Agg CAG GTg ATC AGG TCC CTg ACT GAG aGG CTg TAC
C C D L A P E A R Q V I R S L T E R L Y	781/261	ATT GGC CCC CTG ACC AAC TCC AAg GGC CAG AAC TGT GGC TAC aGg aGG TGC aGg GCC	811/271	ATT GGC CCC CTG ACC AAC TCC AAg GGC CAG AAC TGT GGC TAC aGg aGG TGC aGg GCC
I G G P L T N S K G Q N C G Y R R C R A	841/281	tct GGC GTG CTG ACC ACT AAC TGT GGC AAC ACC CTg ACC TGC TAC CTG AAG GCC TCT GCT	871/291	tct GGC GTG CTG ACC ACT AAC TGT GGC AAC ACC CTg ACC TGC TAC CTG AAG GCC TCT GCT
S G V L T T N C G N T L C Y L K A S A	901/301	Gct TGC aGg GCT GCC AAG CTg CAT GAC TGC ACC ATG CTg GTC TGT GGC GAT GAC CTg GTg	931/311	Gct TGC aGg GCT GCC AAG CTg CAT GAC TGC ACC ATG CTg GTC TGT GGC GAT GAC CTg GTg
A C R A A K L H D C T M L V C G D D L V	961/321	GTg ATC TGT GAg tct Gct GGC ACC CAG GAG GAT GCT GGC tcc CTg aGg GTC TTC ACT GAG	991/331	GTg ATC TGT GAg tct Gct GGC ACC CAG GAG GAT GCT GGC tcc CTg aGg GTC TTC ACT GAG
V I C E S A G T Q E D A S L R V F T E	1021/341	GCC ATG ACC AGG TAC TCT GCC CCC CCT GGC GAC CCT CCC CAg CCT GAG TAT GAC CTG GAG	1051/351	GCC ATG ACC AGG TAC TCT GCC CCC CCT GGC GAC CCT CCC CAg CCT GAG TAT GAC CTG GAG
A M T R Y S A P G D P Q E Y D L E				

FIG. 13B

21/22

1081/361 cTg ATC ACC TCC TGC TCC TCC AAt GTC TCT GTg GCC CAT GAT GCC TCT GGC AAG aGG GTC  
 L I T S C S S N V S V A H D A S G K R V  
 1141/381 1171/391

TAC TAC CTg ACC aGG GAC CCC ACC CCC CTg GCC AGG GCT GCC TGG GAG ACT GCC AGg  
 Y Y L T R D P T T P L A R A A W E T A R

1201/401 1231/411

CAC ACC CCT GTg AAC TCC TGG CTg GGC AAC ATC ATC ATG TAT GCC CCC ACC CTG TGG GCC  
 H T P V N S W L G N I M Y A P T L W A

1261/421 1291/431

AGG ATG ATC CTG ATG ACC CAC TTC TTC TCC ATC CTg CTg GCC CAG GAG CAG CTg GAG AAg  
 R M I L M T H F F S I L A Q E Q L E K

1321/441 1351/451

GCC CTG GGC TGC CAG ATT TAT GGC GCC ACC TAC TTC ATT GAG CCC CTg GAC CTg CCC CAG  
 A L G C Q I Y G A T Y F I E P L D L P Q

1381/461 1411/471

ATC ATC CAG aGG CTg CAT GGC CTg tct GGC TTC TCC CTg CAC tcc TAC TCC CCT GGC GAG  
 I I Q R L H G L S A F S L H S Y S P G E

1441/481 1471/491

ATC AAC AGG GTG GCC TCC TGC CTg AGG AAg CTg GGC GTg CCC CCC CTG aGG GTg TGG AGg  
 I N R V A S C L R K L G V P P L R V W R

1501/501 1531/511

GAC aGG GCC AGG tct GTg aGG GCC AAG CTg CTG TCC CAG GGC AGG GCT GCC ACC TGT  
 H R A R S V R A K L S Q G G R A A T C

1561/521 1591/531

GGC AAG TAC CTg TTC AAC TGG Gct GTG AGG ACC AAG CTg ACC CCC ATC CCT GCT  
 G K Y L F N W A V R T K L T P I P A

FIG. 13C

1621/541	GCC TCC CAG CTG GAC	CTg TCT GGC TGG Ttt	GTg GCT GGC TAC tct	GGC GAC ATC TAC	1651/551
A S Q L D	L S G W F	V A G Y S G	G D I Y		
1681/561	CAC tcc CTG TCC aGg	GCC aGg CCC aGg	GGG TTC ATG TGG TGC	CTg CTg TCT GTg	1711/571
H S L S R A R	R P R W	W C L L L	S V		
1741	GGC GTg GGC ATC TAC	CTG CTg CCC AAC	GGG TGA		1771/591
G V G I Y L	L P N R	*			

### FIG. 13D

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/09884

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : A61N 43/04; C12Q 1/68; C12N 15/00; C07H 21/02; A61K 39/00  
 US CL : 514/44; 435/6, 320.1; 536/23.1; 434/184.1, 192.1

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 514/44; 435/6, 320.1; 536/23.1; 434/184.1, 192.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
none

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, STN, MEDLINE, BIOSIS, EMBASE, CAPLUS, WPIDS, SCISEARCH

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Selby et al. Expression, identification and subcellular localization of the proteins encoded by the hepatitis C viral genome. Journal of General Virology. 1993. Vol. 74, pages 1103-1113, see entire document.	1-3
X	Bukh et al. Sequence analysis of the core gene of 14 hepatitis C virus genotypes. Proc. Natl. Acad. Sci. August 1994. Vol. 91, pages 8239-8243, see entire document.	1-3
Y	Lathe. Synthetic Oligonucleotide Probes Deduced from Amino Acid Sequence Data Theoretical and Practical Considerations. J. Mol. Biol. 1985. Vol. 183, pages 1-12, see entire document.	1-3

Further documents are listed in the continuation of Box C.  See patent family annex.

* Special categories of cited documents:	T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*B* earlier document published on or after the international filing date	Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)		
*O* document referring to an oral disclosure, use, exhibition or other means	A*	document member of the same patent family
*T* document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search  
28 AUGUST 1997

Date of mailing of the international search report

11 SEP 1997

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231  
Facsimile No. (703) 305-3230Authorized officer *Andrew Wang*  
ANDREW WANG  
Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/09884

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Grantham et al. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Research. 1981. Vol. 9, No. 1, pages r43-r74, see entire document.	1-3
A, P	Ide et al. Characterization of the nuclear localization signal and subcellular distribution of hepatitis C virus nonstructural protein NSSA. Gene. December 1996. Vol. 182, pages 203-211, see entire document.	1-3, 8-26
X	US 5,514,539 A (BUKH et al.) 07 May 1996, see entire document.	1-3, 8-26

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/09884

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
  
2.  Claims Nos.: 4-7 because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:  

The enclosed copy of claims 4, 5 were not legible and claims 6, 7 depend on those claims.
  
3.  Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.  As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

The additional search fees were accompanied by the applicant's protest.  
 No protest accompanied the payment of additional search fees.